6 Castrop, J., Hoevenagel, R., Young, J.R. and Clevers, H.C. (1992) *Eur. J. Immunol.* 22, 1327–1330
7 van de Wetering, M. *et al.* (1997) *Cell* 88, 789–799
8 Brunner, E., Peter, O., Schweizer, L. and Basler, K. (1997) *Nature* 385, 829–833
9 Molenaar, M. *et al.* (1996) *Cell* 86, 391–399
10 Lin, R., Thompson, S. and Priess, J.R. (1995) *Cell* 83, 599–609
11 Oosterwegel, M. *et al.* (1993) *Development* 118, 439–448
12 Verbeek, S. *et al.* (1995) *Nature* 374, 70–74
13 Van Genderen, C. *et al.* (1994) *Genes Dev.* 8, 2691–2704
14 van de Wetering, M. and Clevers, H. (1992) *EMBO J.* 11, 3039–3044
15 Giese, K. and Grosschedl, R. (1992) *Cell* 69, 185–195
16 van Houte, L. *et al.* (1995) *J. Biol. Chem.* 270, 30516–30524
17 Love, J.J. *et al.* (1995) *Nature* 376, 791–795
18 Werner, H.M., Huth, J.R., Gronenborn, A.M. and Clore, G.M. (1995) *Cell* 81, 705–714
19 Laudet, V., Stehelin, D. and Clevers, H. (1993) *Nucleic Acids Res.* 21, 2493–2501
20 Giese, K., Kingsley, C., Kirshner, J.R. and Grosschedl, R. (1995) *Genes Dev.* 9, 995–1008
21 Carlson, P., Waterman, M. and Jones, K. (1993) *Genes Dev.* 7, 2418–2430
22 Giese, K. and Grosschedl, R. (1993) *EMBO J.* 12, 4667–4676
23 Bruhn, L., Munnerlyn, A. and Grosschedl, R. (1997) *Genes Dev.* 11, 640–653
24 van de Wetering, M., Castrop, J., Korinek, V. and Clevers, H. (1996) *Mol. Cell. Biol.* 16, 745–752
25 Orsulic, S. and Peifer, M. (1996) *Curr. Biol.* 6, 1363–1367
26 Bhanot, P. *et al.* (1996) *Nature* 382, 225–230
27 Yanagawa, S. *et al.* (1995) *Genes Dev.* 9, 1087–1097
28 Cook, D. *et al.* (1996) *EMBO J.* 15, 4526–4536
29 Orsulic, S. and Peifer, M. (1996) *J. Cell Biol.* 134, 1283–1300
30 Miller, J.R. and Moon, R.T. (1996) *Genes Dev.* 10, 2527–2539
31 Moon, R.T., Brown, J.D., Yang-Snyder, J.A. and Miller, J.R. (1997) *Cell* 88, 725–728
32 Behrens, J. *et al.* (1996) *Nature* 382, 638–642
33 Huber, O. *et al.* (1996) *Mech. Dev.* 59, 3–10
34 Riese, J. *et al.* (1997) *Cell* 88, 777–787
35 Kinzler, K.W. and Vogelstein, B. (1996) *Cell* 87, 159–170
36 Rubinfeld, B. *et al.* (1996) *Science* 272, 1023–1026
37 Moon, R.T. and Miller, J.R. (1997) *Trends Genet.* 13, 256–258
38 Korinek, V. *et al.* (1997) *Science* 275, 1784–1787
39 Morin, P.J. *et al.* (1997) *Science* 275, 1787–1790
40 Rubinfeld, B. *et al.* (1997) *Science* 275, 1790–1792
41 Yost, C. *et al.* (1996) *Genes Dev.* 10, 1443–1454
42 Pai, L.M., Orsulic, S., Bejsovec, A. and Peifer, M. (1997) *Development* 124, 2255–2266

H. Clevers *and* ***M. van de Wetering*** *are in the Department of Immunology, University Hospital Utrecht, PO Box 85500, 3508GA, Utrecht, The Netherlands.*

The case for epigenetic effects on centromere identity and function

GARY H. KARPEN (karpen@salk.edu)

ROBIN C. ALLSHIRE (robin.allshire@hgu.mrc.ac.uk)

The centromere is required to ensure the equal distribution of replicated chromosomes to daughter nuclei. Centromeres are frequently associated with heterochromatin, an enigmatic nuclear component that causes the epigenetic transcriptional repression of nearby marker genes (position-effect variegation or silencing). The process of chromosome segregation by movement along microtubules to spindle poles is highly conserved, yet the putative cis-acting centromeric DNA sequences bear little or no similarity across species. Recently, studies in several systems have revealed that the centromere itself might be epigenetically regulated and that the higher-order structure of the underlying heterochromatin contributes to centromere function and kinetochore assembly.

'I confess that I cannot recall any case within my experience which looked at the first glance so simple, and yet which presented such difficulties.'
The Man with the Twisted Lip[1].

'Chance has put in our way a most singular and whimsical problem, and its solution is its own reward.'
The Adventure of the Blue Carbuncle[1].

The centromere is cytologically visible as the primary constriction of the chromosome, and is associated with the trilaminar kinetochore, which serves as the key site of attachment between the chromosome and the spindle during mitosis and meiosis[2]. Centromere function is essential for mediating chromosome movement along microtubules, and also for normal sister-chromatid cohesion and separation[3]. In this review we attempt to bring together information about the structure and function of the centromere in different eukaryotic systems; we emphasize the emerging, surprising view that epigenetic mechanisms might play a key role in determining centromere identity and regulation. The term 'epigenetic' has acquired different meanings in different contexts. Here, we use epigenetic to mean heritable changes in gene or centromere activity without a corresponding change in primary DNA sequence.

Centromeric DNA sequence is not conserved in different eukaryotes

A mere decade ago it seemed so elementary. The *cis*-acting DNA sequences required to confer complete centromere function had been dissected in great detail in the unicellular budding yeast *Saccharomyces cerevisiae*. The *S. cerevisiae* minimal centromere (CEN) is only 125 bp in length, and all 17 centromeres contain the same three functionally distinct elements (CDEs I, II and III) (Fig. 1). Thus, centromere function in this yeast is known to be mediated by specific DNA sequences, some of which

 0168-9525/97/$17.00
PII: S0168-9525(97)01298-5

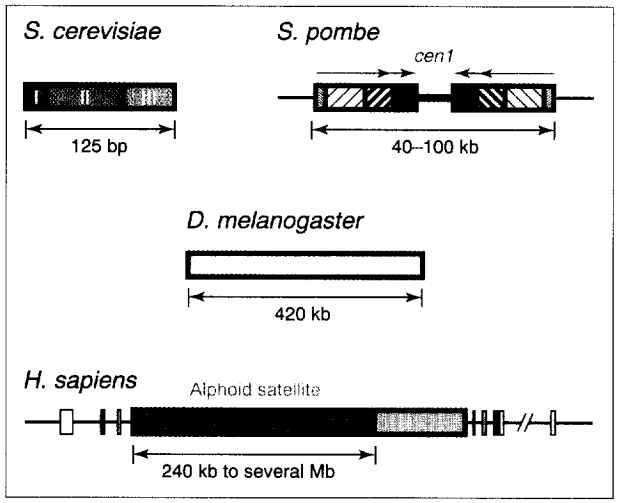

FIGURE 1. Structures of centromeric DNAs in eukaryotic organisms. *S. cerevisiae*: only 125 bp is required for centromere activity. CDEIII is the essential 'business end' of the centromere, the only region whose primary sequence must be conserved for normal function. CDEII elements are very AT-rich, but the exact sequence is not conserved among the different *S. cerevisiae* centromeres; nevertheless, CDEII is required for normal stability and the maintenance of sister chromatid cohesion[65]. CDEI is required for full centromere function, but a complete deletion causes only a 10-fold decrease in stability; CDEI might act as an enhancer of CDEII and III (reviewed in Refs 4, 5). *S. pombe*: the three centromeres have a central core (thick purple line) surrounded by inner repeats (magenta arrow head), collectively termed the central domain, and outer repeats (green arrows). The arrangement and number of repeats differ at each centromere. The simplest centromere *cen1* is shown and occupies approximately 40 kb, *cen2* and *cen3* occupy 65 and 97 kb, respectively, in the standard wild-type strain 972 (reviewed in Ref. 66). *D. melanogaster*: the centromere of the X chromosome, as defined by analyses of irradiation derivatives of the *Dp1187* minichromosome, occupies 420 kb and contains satellite arrays and several complete or partial copies of various transposable elements[8,9]. *H. sapiens*: structure of the Y chromosome centromere. Arrays of alphoid satellite (olive green) ranging from 240 kb to several megabases are found at all conventional human centromeres. The basic alphoid-repeat unit is an AT-rich 171 bp tandem repeat; alphoids present on chromosomes other than the Y usually contain the 17 bp CENPB-binding motif. Other types of satellite repeats (different grey and white boxes) flank the alphoid array e.g. satellite III (Refs 13, 14).

have been demonstrated to contain unique motifs that bind essential centromere proteins[4,5].

It was comforting to imagine that the larger centromeres in other eukaryotes were simply multimeric versions of the *S. cerevisiae* centromere, formally, or even in actual sequence. Despite significant progress over the past decade, we still have an incomplete understanding of the structure and function of the centromere in the fission yeast *Schizosaccharomyces pombe*, and in the multicellular eukaryotes *Drosophila*, mouse and human. Nevertheless, we do know that the centromeres in these organisms are quite different from those in *S. cerevisiae*: they require significantly more DNA for centromere function, and the primary sequence of the centromere is not conserved among species, or even among different chromosomes of an individual organism (Fig. 1).

S. pombe has a similar genome size to *S. cerevisiae*, yet its centromeric domains are much larger and more complex. The three *S. pombe* centromeres contain a 4–7 kb single-copy central core and 20–100 kb of flanking repetitive sequences, arranged as long inverted structures (Fig. 1). Reasonable centromere function is observed for constructs that carry one central core plus inner repeats (the central domain) and two flanking outer repeats. Centromere sequences are significantly less conserved among different *S. pombe* chromosomes, in contrast to *S. cerevisiae*[6]. The central core of centromere 2 lacks a 3.3 kb sequence shared by centromeres 1 and 3 (Ref. 7), and the flanking elements present within a few kilobases of the core (the inner repeats) are unique to each centromere. While the outer repeats share a high degree of conservation among centromeres, they are organized differently. In addition, sequences can be inserted into and deleted from various regions of the *S. pombe* centromere without affecting stability, suggesting that the requirements for centromere function are more malleable in *S. pombe* than in *S. cerevisiae*.

Until recently, the presence of highly repeated sequences has greatly inhibited the analysis of centromeric DNA in multicellular eukaryotes. Molecular-genetic analyses demonstrate that a *Drosophila* minichromosome (*Dp1187*) centromere is contained within a specific 420 kb region of the centric heterochromatin (Fig. 1)[8,9]. Two different simple repetitive sequences comprise most (~350 kb) of the functional centromere, while the remainder consists of interspersed single, complete transposons and a novel AT-rich sequence. Unexpectedly, these satellite DNAs and transposable elements are found at many chromosomal sites that do not form centromeres, and are not present at all centromeres[9]. In addition, centromere function is developmentally regulated in *Drosophila*; the amount of this 420 kb region required for normal chromosome transmission differs in different types of divisions[8]. These observations together suggest that centromere formation and function in *Drosophila* might not depend solely on primary DNA sequence. However, the structural analysis of this centromere is not complete to the DNA sequence level, and it is possible that a very short 'magic sequence' is present, which nucleates the assembly of the 420 kb region into a functional centromere.

Numerous studies have associated large arrays (>500 kb) of tandemly repeated satellite DNAs with centromeres in mammals (Fig. 1), but identification of a specific role for satellite DNA in kinetochore formation or function has been lacking[10]. Elegant analyses of rearranged mammalian chromosomes indicate that the presence of alphoid DNA (an AT-rich, ~171 bp tandem repeat[11]) is correlated with chromosome stability[12–15]. However, alphoid DNA inserted into ectopic (non-centromeric) chromosome sites does not form a functional kinetochore, as judged by the absence of the kinetochore component CENPC (Refs 16–19). Furthermore, centromeres can be inactivated in stable dicentrics, despite the presence of alphoid DNA (see below). The exact function of satellite DNAs at mammalian centromeres is unclear because the transmission behaviour of molecularly defined components has not been assayed directly. Recent transfection studies with purified alphoid DNA have led to the recovery of extrachromosomal

elements; however, centromere function cannot be unequivocally attributed to the alphoid DNA in these studies, because significant structural rearrangements occurred during the production of the stable elements, and their structure and composition are unknown[20,21].

In summary, there are many gaps in our understanding of the structure of functional centromeres in most eukaryotes. Existing data suggest that the *S. cerevisiae* model for centromere structure is not directly applicable to *S. pombe* or multicellular eukaryotes. There is no observed sequence conservation among different species (Fig. 1); the only common theme apparent, to date, is the presence of AT-rich DNA. Indeed, specific centromeric elements are not necessarily conserved among the different chromosomes in the same cell in *S. pombe*, *Drosophila* and mammals. These observations suggest that kinetochore assembly at a particular site on a chromosome might not be dependent on a specific primary DNA sequence and, instead, might be regulated by combinations of different sequences, or by the adoption of a particular higher-order DNA–protein structure. While it is theoretically possible that a small sequence, conserved among all centromeres within each species, or even among different species, remains to be discovered, we think this is unlikely. However, more refined structural and functional analyses in flies and mammals are needed to test this 'magic sequence' hypothesis directly.

Centromeric DNA is not always competent to assemble a functional kinetochore

The mere presence of centromeric DNA does not ensure kinetochore formation and function. For example, the mouse kinetochore is confined to a specific subregion of a much larger, uniform array of minor satellites[22]. Similarly, the major satellite and transposon components of the *Drosophila Dp1187* centromere are present predominantly in regions that are never associated with centromere activity[8,9]. An elegant study demonstrates that kinetochore assembly in *S. pombe* is epigenetically regulated[23]. When a truncated centromere construct, consisting of one central core plus one outer repeat, was transformed into *S. pombe* cells, a functional centromere was only established in a low and variable proportion of transformants (Fig. 2a). Transformants having centromere function were observed to switch to the centromere active state at a low frequency (0.6% of cells). Once a functional centromere is assembled, the active state is perpetuated within a lineage for many divisions[23]. Thus, the same DNA sequences can have two functionally different states.

Centromeres can also be inactivated with no change in DNA sequence. Normally, the presence of more than one centromere on a chromosome (a dicentric) results in the formation of an anaphase bridge, which leads to chromosome loss. However, dicentric chromosomes have been recovered in flies and mammals that appear to be transmitted normally[24,25] (Fig. 2b). In most stable human dicentric chromosomes only one centromere is active, as judged by the presence of essential kinetochore components[16,26–29]. In some cell lines, the functional state of the two centromeres might alternate within a population[29], suggesting the existence of a 'plastic' centromere inactivation mechanism. It would be interesting to determine whether the cell senses the presence of a dicentric chromosome and induces the inactivation of one centromere.

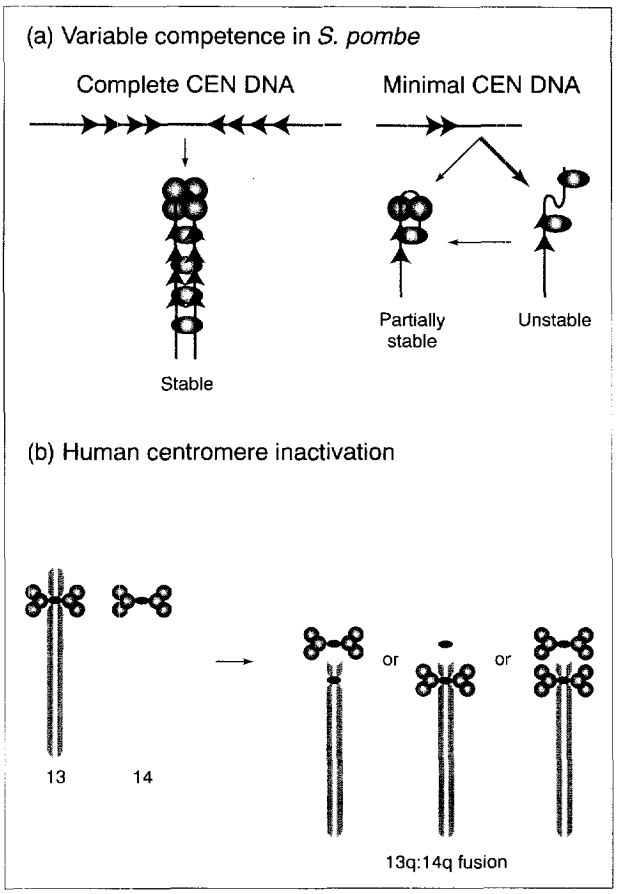

FIGURE 2. Normal centromere-associated DNA alone does not ensure centromere function. (a) In *S. pombe*, plasmids containing a central domain and inverted flanking elements form a functional centromere efficiently when introduced as naked DNA. Gold circles indicate the presence of functional kinetochore components, which are assembled only in the presence of a hypothetical higher-order structure that depends on protein-mediated interactions (grey ovals) between the flanking repeats. Truncated centromeric constructs (e.g. central domain plus one flanking subrepeat) form two types of transformants, which contain the same unrearranged DNA. In most transformants, the plasmid is highly unstable and centromere activity is absent, but occasionally the construct is stably propagated, owing to the assembly of an active centromere[23]. See Fig. 1 legend for other symbols. (b) Fusion of two human chromosomes, such as chromosome 13 and 14, leads to the formation of a dicentric where one or the other centromere can be inactivated. Functional kinetochore components (gold circles) can be associated with the 13 or 14 alphoid-containing regions (purple ovals), suggesting that centromere activation/inactivation might oscillate frequently[29].

Variable competence of centromeric DNA and centromere inactivation argue strongly that what is normally considered as centromeric DNA is not always sufficient to nucleate a functional kinetochore, and that some other parameter must determine the choice and activity of centromere assembly sites.

Noncentromeric DNA can acquire centromere function

The case for kinetochore assembly being independent of a particular *cis*-acting DNA sequence is further supported by examples where centromere function is

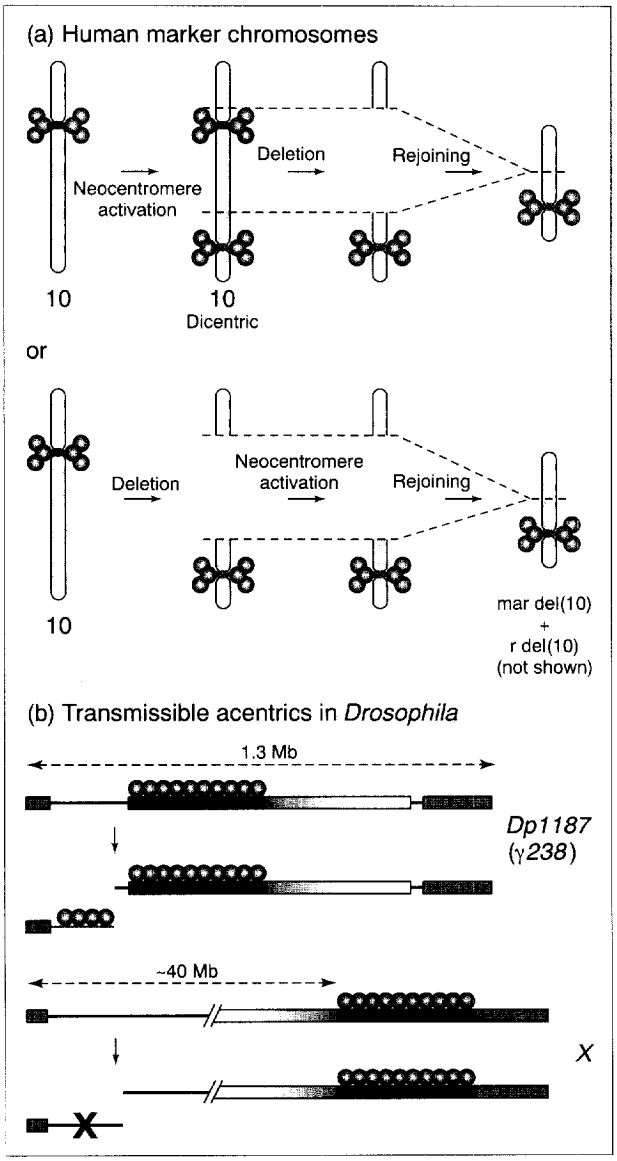

FIGURE 3. Active centromeres can be formed at sites that do not contain sequences ordinarily associated with a centromere. (a) There are many examples of human 'neocentromere' formation in the literature. In one well-studied case [mar del(10)], kinetochore proteins and centromere function (gold circles) reside at a site (10q25) (orange ovals) that lacks alphoid repeats (purple ovals) or other centromere-associated sequences. mar del(10) and the ring chromosome r del(10) (which contains the normal chromosome 10 centromere and alphoid array) might have arisen from activation of 10q25 in a normal chromosome, followed by breakage and rejoining (dashed lines, top). Alternatively, deletion might have preceded neocentromere activation (bottom). The nature of the initial centromere activation event is unknown but mapping across the mar del(10) neocentromere suggests that the DNA is unremarkable and is not homologous to known centromeric DNA (Ref. 34). (b) The *Drosophila Dp1187* minichromosome inversion derivative *γ238* contains euchromatin juxtaposed with the fully functional centromeric DNA (purple). Irradiation mutagenesis produces a high frequency of structurally acentric fragments (orange line) that are transmitted surprisingly well through mitosis and meiosis[8], sharing many properties and components with endogenous *Drosophila* centromeres (gold circles). The importance of centromere proximity to neocentromere activation is suggested by the failure (denoted by a red cross) to recover transmissible acentric fragments from the equivalent region on a complete X chromosome, which in this case lies 40 Mb from the normal X centromere[69].

exhibited by noncentromeric DNA. Rearranged human 'marker chromosomes' (Fig. 3a) have been identified that contain no detectable alphoid-containing regions[30–35]. Nonetheless, a primary constriction is visible that stains with antibodies raised against essential kinetochore components. Although such 'neocentromeric' activity arises rarely in mammals, the recovery and transmission of marker chromosomes demonstrates that normally euchromatic sequences can assemble a kinetochore.

Similar results have been obtained recently in an experimentally manipulable system. Acentric fragments produced by irradiation or laser ablation lack a centromere and are normally lost during cell division in a wide variety of organisms[36–38]. Surprisingly, structurally acentric derivatives of the *Drosophila Dp1187* minichromosome were recovered after irradiation[8] (Fig. 3b). These acentrics lack the collection of satellites and transposable elements associated with the *Dp1187* centromere, and are composed of only 225–290 kb of subtelomeric heterochromatin and euchromatin. Nevertheless, these acentric derivatives are transmitted reasonably well from males (~5% loss per division, on average), although they are transmitted very poorly from females[8]. Three additional lines of evidence strongly suggest that *Dp1187* acentric derivatives are transmitted by a microtubule-based mechanism and exhibit 'neocentromere' activity. Acentric fragments bind the centromere marker protein ZW10 in male meiosis, are often associated with the spindle pole in male meiotic anaphases, and are stabilized by antipoleward forces generated by the NOD kinesin-like protein, all properties of normal centromere-containing chromosomes[69].

How can noncentromeric DNA acquire and propagate centromere function in *Drosophila* and humans? The *Drosophila* minichromosome acentric fragments are amenable to experimental manipulation because they are recovered *de novo* at a high frequency and in a single step[8]. The region that has acquired neocentromere activity normally resides at the tip of the X chromosome, 40 Mb away from the centromere (Fig. 3b). Recent studies demonstrate that newly created acentric fragments from the X tip do not exhibit neocentromeric activity or localize ZW10 (Ref. 69); the acquisition of centromere function by acentric fragments might require proximity to a normal centromere, the arrangement found specifically in the *γ238* minichromosome (Fig. 3b). Therefore, neocentromere activation in flies and other eukaryotes might occur by the spreading of centromere function or proteins from the 'normal' centromere into adjacent regions, or even *in trans*. Alternatively, neocentromeres could arise when a repressive mechanism is removed; the presence of a normal centromere might generally inhibit neocentromere formation *in cis*. In this context, it is worth noting that there are organisms where different parts of the chromosome elaborate kinetochores in different kinds of cell divisions, during normal development. For example, in nematode gonial-cell divisions, the centromere is not localized but, instead, microtubule-insertion sites are distributed along the entire chromosome (holocentric behaviour)[39,40]; perhaps remnants of holocentric activity are retained by noncentromeric regions in other eukaryotes and are actively repressed.

The continued transmission of human marker chromosomes and *Drosophila* acentric fragments in the absence of 'normal' centromeric DNA argues that, once acquired,

neocentromere function is propagated through replication and division. We propose that propagation probably utilizes an epigenetic marking mechanism, which might also regulate the number and types of active centromeres in normal chromosomes.

Centromeric chromatin structure is unusual

In organisms where no specific DNA sequence appears to be necessary or sufficient for centromere function, centromere identity might be determined by a specific 'higher-order structure,' that is, secondary or tertiary structures inherent to the DNA (e.g. DNA bending) or produced by associations with protein (chromatin structure). Are there unusual higher-order structures at different eukaryotic centromeres, and do they share any common features? An obvious indicator of a higher-order structure at eukaryotic centromeres is the primary constriction, which characterizes active conventional centromeres and neocentromeres. Furthermore, in most multicellular eukaryotes the centromere is contained within an enigmatic component of the genome, called 'heterochromatin.' But what is heterochromatin? The term was originally used by Heitz in 1928 to describe regions of moss chromosomes that stained differentially and remained condensed throughout the cell cycle, but it has been adapted and redefined in response to discoveries about its molecular, genetic and biochemical properties (reviewed in Refs 41–43; Box 1). Could heterochromatin structure and behaviour be the key to unlocking the secrets of centromere activity and kinetochore assembly in eukaryotes?

There is biochemical evidence for the presence of a particular type of chromatin structure at yeast and mammalian centromeres. Even in *S. cerevisiae*, a 220–250 bp region of the centromeric DNA is protected from nucleases, and flanking nucleosomes are precisely positioned and phased with respect to this structure[44,45]. The *CSE4* gene of *S. cerevisiae* encodes a variant of histone H3, and *cse4-1* mutations cause elevated rates of centromere dysfunction[46]. Interaction of the *CSE*-H3 variant with histone H4 might form a centromere-specific nucleosome-like particle at CDEII or be incorporated into the surrounding phased nucleosomes to create a unique structure[47].

An unusual chromatin structure also covers a substantial part of the central domain at each *S. pombe* centromere. Although the exact structure of this chromatin is unknown, it appears to lack regularly spaced nucleosomes and is only assembled in the context of a functional centromere[7,48]. A link between this chromatin structure and centromere function has been bolstered by recent studies of Mis6p, which is required for normal centromere function and positioning, and interacts with the central domain. Mutations in *mis6* disrupt the central domain chromatin structure, producing normal nucleosomal arrays[49]. Perhaps the *S. pombe* centromeric chromatin normally incorporates centromere-specific histone variants, or contains histones with altered acetylation patterns. Underacetylation of the N-terminal lysine residues of histones H3 and H4 is a feature of centric heterochromatin in humans[50,51]. Recent studies indicate that chromatin associated with fission yeast centromeres is, in general, underacetylated and that the acetylation state of centromeric histones might play a role in the regulation of centromere activity (K. Ekwall and R.C. Allshire, unpublished).

Box 1. 'Bluffer's guide' to heterochromatin

Heterochromatin is an often used, yet amorphous, term that is difficult to define. Heitz first identified heterochromatin as a differentially stained fraction of nuclear chromatin that is condensed throughout the cell cycle. However, many genomic regions are also associated with heterochromatic properties but are not cytologically distinct (e.g. yeast telomeres). Heterochromatin is generally concentrated in centromeric and telomeric regions of eukaryotic chromosomes, but can be present at interstitial euchromatic sites.

General molecular-genetic properties of heterochromatin have emerged in the past 70 years. Heterochromatin is often found in close association with the nuclear periphery, tends to replicate late in S phase, has a low density of RNA polymerase II transcription units, and has reduced meiotic recombination. The fact that heterochromatin is less accessible to transcription and recombination factors (and artificial probes, such as *dam* methylase) suggests that it consists of a condensed, inaccessible chromatin structure or higher-order structure. Nucleosomes associated with heterochromatin contain histones that are generally underacetylated at N-terminal lysine residues.

An inaccessible chromatin structure should not be confused with a lack of biological activity. Heterochromatic regions are responsible for essential, positive biological functions, including centromere activity, telomere function, nuclear organization, the most highly expressed genes in the genome (ribosomal RNA genes), and the pairing of homologues in meiosis.

Transcriptional repression is exhibited by euchromatic genes placed within or near heterochromatin (or vice versa) owing to chromosomal rearrangements or directed insertions. The repressed/expressed states are subject to epigenetic regulation. Cells assume a repressed or expressed state stochastically; this state is propagated through multiple cell divisions, but can be reversed within a clone/lineage. This form of mosaic transcriptional activity has been termed position-effect variegation (PEV) or reversible transcriptional silencing.

Suppressors and enhancers of variegation [*Su(var)* and *E(var)* mutations] have been identified in *Drosophila* that act to modify the level of repression imposed on a gene subject to PEV. Some of the few *Su(var)* and *E(var)* genes that have been cloned and analysed in detail appear to encode components of heterochromatin, or to regulate its assembly. Chromo- and SET-domain proteins, and histone deacetylases, appear to contribute to heterochromatin formation in many organisms.

Suppression and enhancement of PEV depends on the dosage of certain modifiers, which has led to the idea that heterochromatin assembly is driven by the concentration of available components and might 'spread' or 'ooze' from a nucleation site. However, nuclear positioning also has been demonstrated to affect heterochromatin assembly and gene repression.

Studies of mammalian centromere proteins also suggest that a specific higher-order structure might exist at the centromeres of higher eukaryotes. Dissection of centromere-protein components in mammals has been greatly facilitated by the discovery that antibodies to kinetochore components can be present in the sera of human autoimmune patients (e.g. CENPs A, B, C and F)[2]. CENPA is of particular interest to this discussion because it encodes a variant of histone H3, similar to Cse4p. CENPA-targeting to centromeric regions depends on the histone-fold domain, and on expression late in S phase[52,53]. Centromeric heterochromatin in human cells

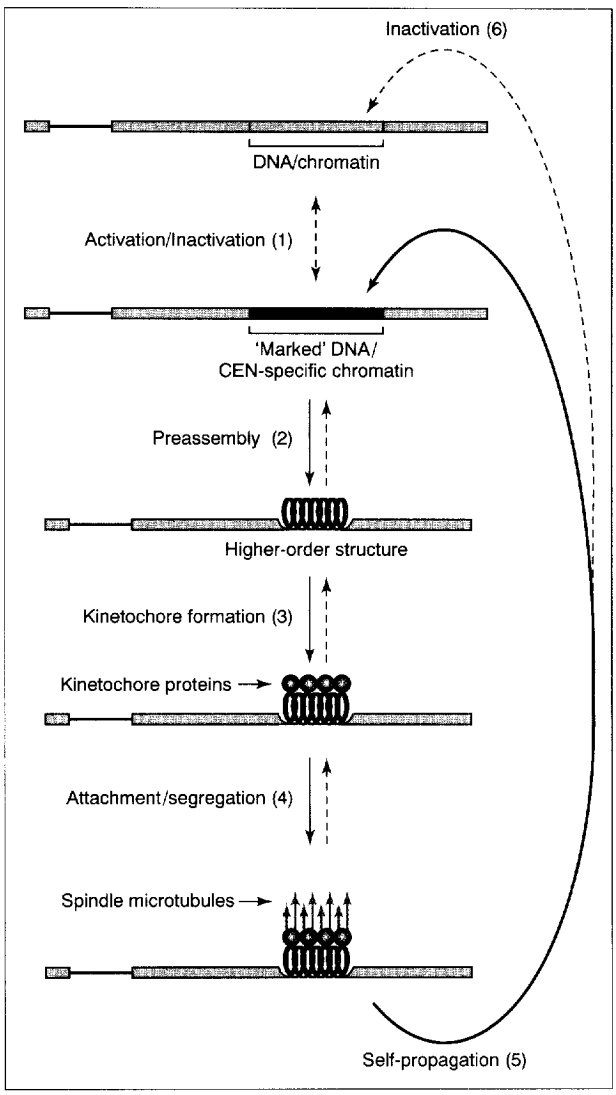

FIGURE 4. Model for self-propagation of centromere identity and function. Only one sister chromatid is shown for simplicity. The grey box is centric heterochromatin (right-hand side) or subtelomeric heterochromatin (left-hand side); the black line is euchromatin; dashed lines indicate less frequent events. (Step 1) Activation/inactivation. Naked centromeric DNA or normally noncentromeric DNA lacks a crucial 'marking' that enables it to be recognized as bona fide centromeric DNA. Alternatively, the ordinary nucleosomes/proteins that assemble on such DNAs might lack a histone variant or a specific pattern of post-translational modification and, thus, fail to assemble CEN-specific chromatin. The types of proteins attracted to centromeric DNA/chromatin might be influenced by its time of replication (see text). Acquisition of these atypical chromatin or DNA characteristics (purple) by an ordinarily euchromatic region could lead to neocentromere activation. This step identifies the region as competent to assemble a centromere-specific higher-order structure. (Step 2) Preassembly of a higher-order structure might occur through self-assembly (e.g. of AT-rich, 'bendable' DNA) or through the binding of new proteins. Reduction in the amount of particular sequences or critical proteins might reduce the efficiency of assembly and, thus, kinetochore nucleation. Neocentromere activation might also occur at the preassembly step; under the right conditions (e.g. proximity to an active centromere *in cis* or *in trans*) unmarked sequences might acquire the centromeric conformation. (Step 3) Kinetochore assembly. Higher-order structure is the foundation for kinetochore assembly (it attracts the key kinetochore assembly proteins; gold circles), rather than primary DNA sequence. (Step 4) Spindle attachment. The assembled kinetochore captures spindle microtubules (purple arrows, heads towards the pole) and segregates chromosomes to daughter nuclei. Antipoleward forces and tension[67] are necessary to maintain attachment to the spindle, and are known to require interaction with extracentromeric regions[68]. (Step 5) Self-propagation. In the next round of replication and division, a kinetochore is assembled at the same site as previous centromere activity, owing to retention of centromere identity by an unidentified marking mechanism (see text). (Step 6) Inactivation. Centromere inactivation could be caused by a self-propagation failure (6), or by reversal of the marking (1) or preassembly (2) steps. Inactivation would remove the centromeric-DNA/chromatin marking, and would render centromeric DNA and neocentromeres unable to assemble a kinetochore and propagate centromere identity, until a new, rare activation event (1) occurred. It is likely that centromere inactivation is also rare, except when special circumstances arise, such as the formation of a stable dicentric chromosome, or on minimal constructs that might be inefficient at assembling and maintaining the active state (Fig. 3). Note that reversal of spindle attachment (4) or kinetochore formation (3) in any one division would not necessarily remove marking and self-propagation of centromere identity, providing a possible basis for the variable expression of centromere activity in stable dicentrics.

appears to be replicated in the later stages of S phase[54]. An attractive model is that late replication leads to the specific incorporation of CENPA at centromeres. However, many other late replicating regions of the genome do not incorporate CENPA, and we do not really know when the functional centromeric DNA itself is replicated, because the precise sequences involved in centromere propagation in humans have not been identified.

Gene silencing links chromatin structure to centromere function

Are specialized centromeric-chromatin structures necessary for normal centromere function? The answer is a clear yes, thanks to genetic studies in *S. pombe* and flies.

One of the classical properties of heterochromatin is the ability to repress gene function, [position-effect variegation (PEV) or reversible transcriptional silencing; see Box 1; reviewed in Refs 41, 43]. Genes placed within fission-yeast centromeres are transcriptionally inactivated in a mosaic fashion: genetically identical cells can exhibit a repressed or expressed marker gene[55] This is very similar to the phenomenon of PEV associated with centric heterochromatin in *Drosophila*.

Mutations that alleviate centromere-induced gene silencing also interfere with chromosome segregation, providing a critical link between chromatin structure and centromere function. Mutations in three (*clr4*, *rik1* and *swi6*) of the six genes required to maintain silencing at the fission-yeast-mating-type loci also alleviate centromere-induced silencing, and cause a 100-fold elevated rate of chromosome loss and a high incidence of lagging chromosomes in late anaphase[56,57]. Twelve additional loci are known to alleviate repression only at centromeric sites and also disrupt chromosome segregation (K. Ekwall and R.C. Allshire, unpublished). These observations suggest that the assembly of a fully functional kinetochore mediates repression at the centromere,

and that some of these gene products might help create the higher-order structure required for normal kinetochore assembly. Swi6p co-localizes with centromeres, telomeres and the silent-mating-type loci, the three known heterochromatic regions in fission yeast nuclei. Localization of Swi6p is disrupted in *clr4* and *rik1* mutants, suggesting that all three products contribute to the assembly of a fully functional kinetochore[56,57].

What types of proteins affect centromere silencing as well as centromere function? Rik1p bears little resemblance to any known proteins or motifs in existing databases, whereas Clr4p and Swi6p contain chromodomains (Ref. 58; P. Lord and R.C. Allshire, unpublished). Intriguingly, chromodomains were originally defined as a region of similarity between the *Drosophila* proteins PC (Polycomb) and HP1 [Heterochromatin protein 1, the product of *Su(var)2-5* (see Box 1)], which are involved in different types of transcriptional repression. Mammalian HP1-like proteins have also been identified that localize to centric heterochromatin[59]. In addition to an N-terminal chromodomain, Clr4p and another fly chromodomain protein, encoded by *Su(var)3-9*, contain C-terminal SET domains[60]. Recent analyses demonstrate that some *Su(var)* genes also affect chromosome segregation in *Drosophila* (Refs 61, 62; K. Cook and G.H. Karpen, unpublished), suggesting that heterochromatin structure and centromere function are also linked in multicellular eukaryotes.

An epigenetic model for centromere function

We have summarized evidence that centromeric DNA is not always sufficient for kinetochore formation in *S. pombe*, flies and mammals. Moreover, sequences other than the endogenous centromeric DNA are capable of functioning as centromeres in flies and mammals, although strong regulatory mechanisms must normally exist to prevent their activity. These observations lead us to propose that normal centromere function is regulated in an epigenetic fashion, and that epigenetic regulation will continue to be an important component, even if a specific centromeric 'magic sequence' is eventually identified.

It is surprising that centromere activity and kinetochore assembly display so much plasticity, because centromeres encode an essential and widely conserved function. It is important to note that the site chosen for centromere activity on an individual chromosome does not change dramatically from one generation to the next, at the resolution of light microscopy[22]. How can centromere assembly be plastic and stable at the same time? We propose a model for normal centromere assembly and function that accounts for these unusual properties (Fig. 4). Briefly, an initial marking event distinguishes centromeric DNA/chromatin from normal bulk DNA/chromatin (1). This atypical DNA/chromatin is converted into a higher-order structure (2), which acts as a nucleation site for kinetochore assembly (3) and activity (4). Centromere identity is propagated throughout division by maintenance of the epigenetic marking (5), and inactivation of a centromere could result from loss of the centromere-specific marking (6).

Simply put, the key feature of this model is the self-propagation of centromere identity and function during normal replication. We propose that a site has centromere function because it was utilized as a centromere in the previous division.

How can the site chosen for centromere activity be self-propagating? The mechanism must be linked to replication to allow the templating of a specific protein or structure onto daughter chromatids. DNA modification is an obvious potential mechanism because it is correlated with other types of epigenetic inheritance, including X inactivation and genomic imprinting[63]. However, *Drosophila* and *S. pombe* lack methylated bases and would have to utilize some currently unknown type of DNA modification. Alternatively, centromere identity might be inherited via a protein-based marking mechanism[50]. For example, centromere-specific histone variants (e.g. CENPA) or histones with unique acetylation patterns could produce a centromere-specific chromatin structure that would be seeded on newly replicated DNA via specific protein–protein interactions. Although it seems outrageous, it is also possible that centromere inheritance is determined by a prion-like protein-templating mechanism[64]. Finally, the propagation of centromere identity could be facilitated by replication late in S phase, when the concentration of key components (such as CENPA) is favourable to the assembly of the CEN-specific complex. Late replication might depend on the constitution of this centromere-specific chromatin, giving rise to an interdependent relationship. The formation and propagation of neocentromeres might be explained by a shift in the replication timing at unusual sites. In *Drosophila*, neocentromeres could arise owing to prior association with bona fide centromeres, suggesting that a centromere might cause adjacent sequences to take on centromeric-marking or replication patterns, just as heterochromatin can influence the expression of juxtaposed genes.

The self-propagation/epigenetic model proposed here can account for both the stability and the plasticity of the centromere. 'Unmarked' sites would only rarely acquire centromere activity, thus ensuring monocentric behaviour in normal chromosomes; but once a site is utilized as a centromere, its activity would be perpetuated. Many new and old tools can now be used to test specific models for epigenetic and genetic determination of centromere identity and function. For example, is neocentromere activation correlated with a switch in replication timing? Can interference with replication timing alter the state of active and inactive centromeres on human dicentric chromosomes and in *S. pombe*? Can targeting of proteins such as Swi6p, CENPA and CENPC to noncentromeric regions induce kinetochore formation? Do alterations in the acetylation pattern of centromeric chromatin disrupt centromere function? This is an exciting time for centromere enthusiasts: the game is afoot!

> *'This is an instructive case. There is neither money nor credit in it, yet one would wish to tidy it up.'*
> The Adventure of the Red Circle[1].

Acknowledgements

We thank members of our laboratories, and other colleagues, for comments on the manuscript, and for sharing information and ideas, which aided the evolution of this review. We apologize to those colleagues whose work we did not include or quote due to space limitations. Centromere research in the authors' laboratories is supported by grants from the American Cancer

Society (DB-1200), NIH (R01s HG00747 and GM54549), and a Pew Fellowship (G.H.K.), and The Medical Research Council of Great Britain (R.C.A.).

References

1 Doyle, A.C. (1968) in *The Annotated Sherlock Holmes* (Baring-Gould, W.S., ed.), J. Murray
2 Pluta, A.F. *et al.* (1995) *Science* 270, 1591–1594
3 Miyazaki, W.Y. and Orr-Weaver, T.L. (1994) *Annu. Rev. Genet.* 28, 167–187
4 Hegemann, J.H. and Fleig, U.N. (1993) *BioEssays* 15, 451–460
5 Hyman, A.A. and Sorger, P.K. (1995) *Annu. Rev. Cell Dev. Biol.* 11, 471–495
6 Takahashi, K. *et al.* (1992) *Mol. Biol. Cell* 3, 819–835
7 Clarke, L. *et al.* (1993) *Cold Spring Harbor Symp. Quant. Biol.* 58, 687–695
8 Murphy, T.D. and Karpen, G.H. (1995) *Cell* 82, 599–609
9 Sun, X., Wahlstrom, J. and Karpen, G.H. *Cell* (in press)
10 Tomkiel, J.E. and Earnshaw, W.C. (1993) in *Chromosome Segregation and Aneuploidy* (Vig, B.K., ed.), pp. 13–29, Springer-Verlag
11 Choo, K.H. *et al.* (1991) *Nucleic Acids Res.* 19, 1179–1182
12 Tyler-Smith, C. *et al.* (1993) *Nat. Genet.* 5, 368–375
13 Brown, K.E. *et al.* (1994) *Hum. Mol. Genet.* 3, 1227–1237
14 Heller, R., Brown, K.E., Burgtorf, C. and Brown, W.R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 7125–7130
15 Farr, C. *et al.* (1995) *EMBO J.* 14, 5444–5454
16 Earnshaw, W.C., Ratrie, H., III, and Stetten, G. (1989) *Chromosoma* 98, 1–12
17 Haaf, T., Warburton, P.E. and Willard, H.F. (1992) *Cell* 70, 681–696
18 Larin, Z., Fricker, M.D. and Tyler-Smith, C. (1994) *Hum. Mol. Genet.* 3, 689–695
19 Warburton, P.E. and Cooke, H.J. (1997) *Chromosoma* 106, 149–159
20 Taylor, S.S., Larin, Z. and Tyler-Smith, C. (1996) *Chromosoma* 105, 70–81
21 Harrington, J.J. *et al.* (1997) *Nat. Genet.* 15, 345–355
22 Mitchell, A. *et al.* (1996) *J. Cell Sci.* 109, 2199–2206
23 Steiner, N.C. and Clarke, L. (1994) *Cell* 79, 865–874
24 Ault, J.G. and Lyttle, T.W. (1988) *Chromosoma* 97, 71–79
25 Therman, E., Sarto, G.E. and Patau, K. (1974) *Am. J. Hum. Genet.* 26, 83–92
26 Merry, D.E., Pathak, S., Hsu, T.C. and Brinkley, B.R. (1985) *Am. J. Hum. Genet.* 37, 425–430
27 Therman, E., Trunca, C., Kuhn, E. and Sarto, G. (1986) *Hum. Genet.* 72, 191–195
28 Page, S.L., Earnshaw, W.C., Choo, K.H. and Shaffer, L.G. (1995) *Hum. Mol. Genet.* 4, 289–294
29 Sullivan, B.A. and Schwartz, S. (1995) *Hum. Mol. Genet.* 4, 2189–2197
30 Voullaire, L.E., Slater, H.R., Petrovic, V. and Choo, K.H. (1993) *Am. J. Hum. Genet.* 52, 1153–1163
31 Ohashi, H. *et al.* (1994) *Am. J. Hum. Genet.* 55, 1202–1208
32 Blennow, E. *et al.* (1994) *Am. J. Hum. Genet.* 54, 877–883
33 Brown, W. and Tyler-Smith, C. (1995) *Trends Genet.* 11, 337–339
34 DuSarte, D. *et al.* (1997) *Nat. Genet.* 16, 144
35 Depinet, T.W. *et al.* (1997) *Hum. Mol. Genet.* 6, 1195–1204
36 Mather, K. and Stone, L.H.A. (1933) *J. Genet.* 28, 1–24
37 Liang, H. *et al.* (1994) *Exp. Cell. Res.* 213, 308–312
38 Khodjakov, A., Cole, R.W., Bajer, A.S. and Rieder, C.L. (1996) *J. Cell Biol.* 132, 1093–1104
39 Pimpinelli, S. and Goday, C. (1989) *Trends Genet.* 5, 310–315
40 Albertson, D.G., Rose, A.M. and Villeneuve, A.M. (1997) in *The Nematode C. elegans* (Vol. II) (Riddle, D.L., Blumenthal, T., Meyer, B.J. and Priess, J.R., eds), pp. 47–78, Cold Spring Harbor Laboratory Press
41 Karpen, G.H. (1994) *Curr. Opin. Genet. Dev.* 4, 281–291
42 Lohe, A.R. and Hilliker, A.J. (1995) *Curr. Opin. Genet. Dev.* 5, 746–755
43 Weiler, K.S. and Wakimoto, B.T. (1995) *Annu. Rev. Genet.* 29, 577–605
44 Saunders, M.J., Yeh, E., Grunstein, M. and Bloom, K. (1990) *Mol. Cell. Biol.* 10, 5721–5727
45 Saunders, M., Fitzgerald-Hayes, M. and Bloom, K. (1988) *Proc. Natl. Acad. Sci. U. S. A.* 85, 175–179
46 Stoler, S., Keith, K.C., Curnick, K.E. and Fitzgerald, H.M. (1995) *Genes Dev.* 9, 573–586
47 Smith, M. *et al.* (1996) *Mol. Cell. Biol.* 16, 1017–1026
48 Polizzi, C. and Clarke, L. (1991) *J. Cell Biol.* 112, 191–201
49 Saitoh, S., Takahashi, K. and Yanagida, M. (1997) *Cell* 90, 131–143
50 Jeppesen, P. and Turner, B. (1993) *Cell* 74, 281–289
51 O'Neill, L. and Turner, B. (1995) *EMBO J.* 14, 3946–3957
52 Shelby, R.D., Vafa, O. and Sullivan, K.F. (1997) *J. Cell Biol.* 136, 501–513
53 Sullivan, K.F., Hechenberger, M. and Masri, K. (1994) *J. Cell Biol.* 127, 581–592
54 Ten, H.K., Gilbert, D., Willard, H. and Cohen, S. (1990) *Mol. Cell. Biol.* 10, 6348–6355
55 Allshire, R.C., Javerzat, J.P., Redhead, N.J. and Cranston, G. (1994) *Cell* 76, 157–169
56 Allshire, R.C. *et al.* (1995) *Genes Dev.* 9, 218–233
57 Ekwall, K. *et al.* (1996) *J. Cell Sci.* 109, 2637–2648
58 Lorentz, A., Ostermann, K., Fleck, O. and Schmidt, H. (1994) *Gene* 143, 139–143
59 Wreggett, K. *et al.* (1994) *Cytogenet. Cell Genet.* 66, 99–103
60 Tschiersch, B. *et al.* (1994) *EMBO J.* 13, 3822–3831
61 Wines, D.R. and Henikoff, S. (1992) *Genetics* 131, 683–691
62 Kellum, R. and Alberts, B.M. (1995) *J. Cell Sci.* 108, 1419–1431
63 Riggs, A. and Pfeifer, G. (1992) *Trends Genet.* 8, 169–174
64 Gabizon, R. and Taraboulos, A. (1997) *Trends Genet.* 13, 264–269
65 Sears, D.D., Hegemann, J.H., Shero, J.H. and Hieter, P. (1995) *Genetics* 139, 1159–1173
66 Allshire, R.C. (1996) in *Epigenetic Mechanisms of Gene Regulation* (Russo, V.E.A., Martienssen, R.A. and Riggs, A.D., eds), pp. 443–466, Cold Spring Harbor Laboratory Press
67 Nicklas, R.B., Ward, S.C. and Gorbsky, G.J. (1995) *J. Cell Biol.* 130, 929–939
68 Murphy, T.D. and Karpen, G.H. (1995) *Cell* 81, 139–148

Reference added in proof

69 Williams, B.C., Murphy, T.D., Goldberg, M.L. and Karpen, G. H. *Nat. Genet.* (in press)

G.H. Karpen *is in the Molecular Biology and Virology Laboratory, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.*
R.C. Allshire *is in the Chromosome Biology Section, MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, UK EH4 2XU.*